

11) Publication number:

0 079 739

A2

12

EUROPEAN PATENT APPLICATION

(21) Application number: 82305926.6

(22) Date of filing: 08.11.82

(51) Int. Cl.3: C 12 N 15/00

C 12 N 1/00, C 12 P 21/02 C 07 H 21/04, C 07 C 103/52 //C12R1/19, C12R1/865

30 Priority: 12.11.81 US 320632

Date of publication of application: 25.05.83 Bulletin 83/21

(BA) Designated Contracting States:
BE CH DE FR GB IT LI NL SE

(1) Applicant: THE UPJOHN COMPANY 301 Henrietta Street Kalamazoo, Michigan 49001(US)

(2) Inventor: Dugaiczyk, Achilles c/o The Upjohn Company 301 Henrietta Street Kalamazoo Michigan 49001(US)

(74) Representative: Perry, Robert Edward et al, GILL JENNINGS & EVERY 53-64 Chancery Lane London WC2A 1HN(GB)

(A) Albumin-based nucleotides, their replication and use, and plasmids for use therein.

(5) The DNA sequence coding for human serum albumin has been isolated and inserted as two fragments into two novel plasmids which can be replicated in *E. coli*. These novel fragments can be joined to provide a unitary DNA sequence which then can be cloned into a suitable host, e.g. *E. coli*, for the expression of human serum albumin (which is used extensively in medical practice in treating shock conditions).

FP 0 079 739 A2

GJE 70/2056/02

ALBUMIN-BASED NUCLEOTIDES, THEIR REPLICATION AND USE, AND PLASMIDS FOR USE THEREIN

This invention relates to nucleotides related to human serum albumin (HSA), their replication and use, and plasmids (and host substances) for use therein.

The gene for serum albumin is regulated in 5 development. On the other hand, serum albumin is synthesised in mammals by the adult liver, and its plateau in adulthood. The embryonic liver and yolk sac, on the other hand, produce predominantly α -fetoprotein, but the synthesis decreases drastically after birth. Recently,

10 Law et al determined the complete sequence of mouse α-fetoprotein mRNA, Nature 291 (1981) 201-205. structure revealed extensive homology to mammalian serum albumin, indicating that the two proteins are encoded in the same gene family. Similar conclusions have been

15 reached from studies on the a-fetoprotein genes of the rat and the mouse; see Jagodzinski et al, Proc. Natl. Acad. Sci. USA, 78 (1981) 3521-3525, and Gorin et al, J. Biol. Chem. 256 (1981) 1954-1959.

The complete nucleotide sequence of human serum 20 mRNA has been determined from recombinant cDNA clones and from a primer-extended cDNA synthesis on the mRNA 2,078 nucleotides, comprises The sequence template. starting upstream of a potential ribosome binding site in the 5'-untranslated region. It contains all the 25 translated codons and extends into the poly(A) at the 3'-terminus. Part of the translated sequence codes for a hydrophobic prepeptide met-lys-trp-val-thr-phe-ile-serleu-leu-phe-leu-phe-ser-ser-ala-tyr-ser, followed by a basic propeptide arg-gly-val-phe-arg-arg. These signal 30 peptides are absent from mature serum albumin and, so far, have not been identified in their nascent state in humans. A remaining 1,755 nucleotides of the translated mRNA sequence code for 585 amino acids which are in agreement, with few exceptions, with the published amino

35 acid data for human serum albumin. The mRNA sequence verifies and refines the repeating homology in the tripledomain structure of the serum albumin molecule.

DETAILED DESCRIPTION OF THE INVENTION

Human serum albumin cDNA is cloned into the PstI site of plasmid pBR322 by the oligo(dG)-oligo(dC) tailing technique. Plasmid DNA was isolated from 97 positive colonies which hybridized to the enriched albumin cDNA probe, and the recombinant plasmid pHA36 was found to contain the largest insert of an albumin cDNA sequence. Its restriction endonuclease map is shown in the drawing, together with a restriction map of the primer-extended plasmid clone pHA206. The latter was obtained in a second transformation experiment after initiating the cDNA synthesis from an internal primer. This primer was a 91 base pairs long DNA fragment, MspI(152)-TaqI(182/3), isolated from pHA36. The two plasmids, pHA36 and pHA206, share 0.15 kb of homologous DNA. Together, they encode the entire sequence for human serum albumin, starting with the CTT codon for leu -10 of the prepeptide and extending into the 3'-untranslated region of poly(A).

Sequence of the Albumin cDNA. The sequence was determined for the most part on both DNA strands to ensure accuracy. All of the restriction sites used to end-label DNA fragments were sequenced across by ²⁰ labeling a neighboring restriction site. The entire nucleotide sequence of the serum albumin mRNA, as determined from the cloned DNA in pHA36, pHA206, and from the primer-extended cDNA at the 5'-terminus of the message, is shown in the following Table 1. The inferred amino acid sequence is also indicated. The mRNA length is 2,078 nucleo-²⁵ tides, of which 38 represent the 5'-untranslated region, 54 identify a prepeptide of 18 amino acids, 18 identify a propeptide of 6 amino acids, 1,755 code for the known 585 amino acids of serum albumin, 189 make up the 3'-untranslated region and 24 are the poly(A) sequence. Nucleotides 5 to 15 (-34 to -24) in the 5'-untranslated region (Table 1) are complementary to a 3'-terminal region of eukaryotic 18S RNA [Azad, A.A. and Deacon, N.J. (1980) Nucl. Acids Res. 8, 4365-4376] and thus could represent a ribosome binding site:

(5')...T
$$T^CT$$
 C T T C T G T......albumin mRNA
(3')...G A G G A A G G C G U C C m_2^6 A m_2^6 A.....18S RNA

The translated portion of the mRNA sequence codes for the signal peptide and the main body of the albumin polypeptide chain. The

4083

signal peptide is composed of a hydrophobic prepeptide of 18 amino acids and a basic propeptide of 6 amino acids (Table 1). Since prepeptides are removed from nascent secretory proteins (like albumin) in the endoplasmic reticulum, they are seen only in vitro in heterologous translation systems. As yet, they have not been found within cells [Judah, J.D. and Quinn, P.S. (1977) FEBS 11th Mtg., Copenhagen 50, 21-29; and Strauss, A.W., Donohue, A.M., Bennett, C.D., Rodkey, J.A. and Alberts, A.W. (1977) Proc. Natl. Acad. Sci. USA 74, 1358-1362]. This is the first report of the presence and the sequence of a prepeptide for human serum albumin. As it is with other secretory proteins, the conversion of proalbumin to albumin takes place in the Golgi vesicles, and the enzyme responsible for this cleavage is probably cathepsin B [Judah, J.D. and Quinn, P.S. (1978) Nature 271, 384-385]. This is also a first report on the sequence of the propeptide for normal human serum albumin.

At the 3'-end of the message, the putative polyadenylation signal sequence, AATAAA, is located 164 nucleotides downstream from the amino acid termination codon TAA and 16 nucleotides upstream from the beginning of the poly(A) sequence. Another characteristic sequence located near the polyadenylation site has been identified by Renoist, et al. [Benoist, C., O'Hare, K., Breathnach, R. and Chambon, P. (1980) Nucl. Acids Res. 8, 127-142]; the concensus sequence from several mRNAs was concluded as TTTTCACTGC. A similar sequence, TTTTCTCTGT, is located 19 nucleotides upstream from the AATAAA hexanucleotide in the human albumin mRNA (Table 1).

TABLE 1

				TABI	LE 1				
	(30)	(170)	(192)	(350)	(440)	(330)	(620)	(710)	(300)
-	AGC	20 173 AAA (50 818 677	80 leu CTT	110 pro CCA	140 try TAT	170 aln cAA	200 cys TGT	230 alu GAA
5	ohe s	phe TTC	phe TT	thr	asn AAC	qlu thr phe leu lys lys tyr leu GAG ACA TTT TTG AAA AAA TAC TTA	169 cys TGC	ain ara leu lys CAG AGA CTC AAG	pro lys ala qlu phe ala CCC AAA GCT GAG TTT GCA
	1eu 1	AAT	o o o	ala GCA	ase GAC	tyr TAC	168 cys TGT	leu CTC	ake TT
	phe leu TTT CTC	glu asn GAA AAT	thr	val GTT	GAT	lys lys tyr AAA AAA TAC	ala phe thr glu GCT TTT ACA GAA	2 A	a)u GAG
	3 =	ala GA		th ACA	1 X	1ys AA	thr	ala CAG	ala SCT
10	-10 leu leu CTT CTT	leu aly alu TTG GGA GAA	glu val GAA GTA	55 TGC 35	h is CAC	leu TTG	phe	ala lys GCC AAA	1ys
	ser 1 TCC (7 DE		75 1ys leu cys AAA TTA TGC	phe leu qin his TTC TTG CAA CAC	a te	ala GCT	၁၂၁	နို့ ည
	lle s ATT 1	asp leu GAT TTG		1ys	leu TTG	th ACA	lys ala AAA GCT	36T	220 ser gin arg phe AGC CAG AGA TTT
	a E	lys AAA	E Y	asp	phe TTC	g cyc	lys AAA	ser TCG	arg AGA
	٥ ي ه	phe TT	\$ ¥	91y	101 6ys TGC	g Ju GAA	tyr TAT	ala GCT	CAG
15	p r o trp val tlu phe TGG GTA ACC TTT	10 ala his arg phe lys GCT CAT CGG TTT AAA	40 GTA	75 thr leu phe gly asp lys leu cys thr val ACC CTT TTT CGA GAC AAA TTA TGC ACA GTT	형목정	130 asn AAT	160 arg AGG	190 qiu qiy lys ala GAA GCG AAG GCT	
	673 70 55	10 ala his arg GCT CAT CGG	his CAT	leu CTT	asn	asa GAC	ala lys CCT AAA	alu aly GAA GGG	arg leu CGC CTG
	1ys	ala GCT	esp GAT	thr	క్లి వ్ర	M SAT			55
٠	-18 p r o Met lys trp val tlu phe lle ser CTCTTCTGTCACCCCACACCCTTTGGCACA ATG AAG TGG GTA ACC TTT ATT TCC	va) GTT	40 phe qlu asp his val lys leu val TTT GAA GAT CAT GTA AAA TTA GTG	leu his CTT CAT	gin glu pro aly arg asn CAA GAA CCT GGG AGA AAT	ala phe GCT TTT	phe ahe TTC FTT	glu leu arg asp GAA CTT CGC GAT	ala OCT
20	ACA	es es	phe TTT		pro CCT	ala GCT	Phe TTC	61e 620	val
20	7665	ser AGT	PTO	lys ser AAA TCA	gin glu CAA GAA	124 cys thr TGC ACT	leu CTT	leu Crr	trp ala TGG GCA
	CCTI	1ys AG	34 Cys TCT		416 CAA		ala pro glu leu leu GCC CCG GAA CTC CTT	glu GAA	160
	ונאפט	ala his GCA CAC	gln gln CAG CAG	asp	ala lys GCA AAA	val met GTG ATG	gy gy	gse :	8 8 8 1 GCA
))	a1a GCA				val	pro CCC		1 Jys
25	īcae.	-1 1 arg asp CGA GAT	30 tyr leu TAT CTT	60 glu asn GAA AAT	91 Cy3	120 glu val asp val met GAG GTT GAT GTG ATG		177 ala cys leu leu pro lys leu asp GCC IGC CTG TTG CCA AAG CTC GAT	210 ala phe lys ala trp ala val GCT TTC AAA GCA TGG GCA GTA
	וכדמו	- F 53	30 tyr TAT		80 673 760	120 val	150 tyr	180 J pro 5 CCA	210 9 ala A GCT
	STCT	arg CGT	gln CAG	ala GCT	asb CAC	ule o	phe TTT	1 E	7 AG
	GCTTT10	pro glyvalphe GGT GTG TTT	ala CCT	glu ser GAG TCA	met ala ATG GCT	0 V	tyr TAC	s le	6.9
	ຮ	val r	phe TTT	56.58	glu met CAA ATG	81.9 ACA	s pro	177 177 100 100	- 6 5 L
30		-1 -6 p r o ser arg gly val phe TCC AGG GGT GTG TTT	هاء 200	asp	g Ay	leu val TTG GTG	arg arg his pro ACA ACA CAI CCI		s TP
		arg AGG	11e ATT	ala GCT	419	leu TTG	arg arg AGA AGA	lys ala AAA GCT	<u>`</u> `\$
		-1 ser 100	15	53 cys val ICT CTT	tyr TAT	arg CGA	£ 5		16 Y
		tyr	21 ala leu val leu lle ala GCC TTG GTG TTG ATT GCC	53 cys TGT	glu thr tyr gly GAA ACC TAT GGT	leu pro arg CTC CCC CGA	ala T GCC	a asp T GAT	210 ser leu gin lys phe gly glu arg ala AGT CTC CAA AAA TTT GGA GAA AGA GCT
35		8 g	leu 11G	A S			11e	a ala r ccr	
JJ		100 100	21 ala GCC	2 2 8 8	81 819 CCT	111 asn AAC	141 910 8A9	171 ala cc	201 ala GCC

	(890)	(980)	320 ala מרד (1070)	(1160)	(1250)	(1340)	(1430)	(1520)	(1610)	(1700)
	260 1eu CTT (290 11e ATT (320 ala (xr T	350 81 a GCC	380 leu CCT	ara CGT	840 his CAT	&70 Ser AGT	SOO 1ys	530 val GTT
5		289 2 cys TGC /	tyr		pro CCT		1ys AAA	va) GTA	و ر و (زر	CTT
	arq ala asp AGG GCG CAC	h is CAC	AAC	PTG AGA	lys AAA	leu TTA	438 cys TGT	978 CCA	val	ala GCA
	arg a	ser his TCC CAC	1ys	leu CTG	phe TTT		437 cys TGT	thr	tyr TAC	thr
	asp a	1 ys	316 cys lys TGC AAA	ser vál val leu leu leu ara leu TCT GTC GTG CTG CTG AGA CTT	a Jr GAA		1ys AAA	1 ys	glu thr GAA ACA	520 ser glu lys glu arg gln fle lys lys gln thr TCT GAG AAG GAA CAA ATC AAG AAA CAA ACT
10	ala asp asp CCT CAT CAC	280 glu lys pro leu leu glu lys GAA AAA CCT CTG TTG GAA AAA	val	leu CTG	asp GAT	aln asn CAG AAT	AGC		ole GAA	1ys AAA
	ala CCT	Jet TTG	asb val CAT GTT	val leu GTG CTG	phe TTC	a ta CAG	aly GGC	leu his TTG CAT	asp CAT	1ys AAG
		leu CTG	1ys AAG	vál GTC	val GTG	phe TTC	lys val AAA GTG	leu 11G	490 ala leu glu val GCT CTG GAA GTC	11e ATC
	253 glu cys GAA TGT	970 CCT	310 val glu ser lys GTT GAA AGT AAG	367 TCT	1ys AA	lys AA	lys AA	va] GTG	9 to	520 qlu lys glu arg gln GAG AAG GAG AGA CAA
		1ys	glu ser GAA AGT	340 8sp tyr GAT TAC	313 GCC	400 alu tyr GAG TAC	gg 4	460 461 leu cys TTA TGT	Jeu CTG	AG 4
15	250 Jeu CTG	280 q1u GAA	310 val GTT		370 tyr TAT		430 ser arg asn leu qly TCA AGA AAC CTA GGA	460 461 gln leu cys CAG TTA TGT	490 ser ala TCA GCT	520 1 q1u 1 CAG
	asp CAT		asp phe CAT TTT	pro CCT	369 alu cys GAA TGC	alu aln leu aly GAG CAG CTT GGA	AAC	g S		1 1y3
	÷ 55	278 cys TGC	asp CAT	h1s CAT	a]u GAA	CTT	arg AGA	leu ser val val leu asn CTA TCC GTG GTC CTG AAC	pro cys phe CCA TGC TTT	- 6 ·
	his CAT	a)u GA	ala GCT	arg AGG	pro his CCT CAT	41° CAG	ser TCA	leu CTG	cys 100	ser TCT
	246 cys TGC	1ys AAG	ala GCT	AGA	pro CCT	phe alu TTT CAG	glu val CAG GTC	val	Dro CCA	514 11e cys thr 1eu ATA TGC ACA CTT
20	245 246 glu cys cys GAA TGC TGC	leu CTG	leu TTA	ala GCA	ala ala ala asp GCC GCT GCA GAT	phe TT	95	val GTG	879 CCA	thr ACA
	gle GA	lys AAA	ser	glu tyr GAA TAT	a1 a	glu leu GAG CTT	leu val CTT GTA	3er 100	3 arg	514 11e cys ATA TGC
	his thr CAC ACG	ser	pro CCT		ala GCT	alu GAG	CTT CTT	. leu	asu AAC	4 1
		ser TCC	leu TTG	tyr IAT		392 cys	thr ACT	tyr TAT	val 5 CTG	ala asp GCA GAT
25	240 1ys val AAA GTC	270 ser 11e TCG ATC	300 ala asp leu pro GCT GAC TTG CCT	330 phe leu tyr TTT TTG TAT	360 361 cys cys TCC TCT	390 gln asn CAA AAT	pro CCA	450 ala glu asp GCA GAC) 160	s ala
	240 1ys AAA		300 als GCT	330 phe TTT	360 cys	390 91n	420 ser thr TCA ACT	450 1 glu	480 J Ser A TCC	510 phe h1s TTC CAT
	leu thr CTT ACC	glu asn gin asp GAA AAT CAA GAT	pro CCT	met. ATG	1ys	¥			r 91u A GAA	_
	asp leu GAT CTT	glu asn gln GAA AAT CAA	met ATG	91y	alu GAG	11 ATC	val CTG	448 cys : TGT	476 477 cys cys thr TGC TGC ACA	A G
	asp GAT	asn	glu GAG	leu TTG	leu CTA	1 le	oll CAA	, S	476 477 cys cys TCC TCC	Å L
30		g &	asp GAT	phe TTC	thr.	. AA1	,) ne	476 s cys A TCC	a ₹
	val 616	265 11e cys ATC 1GT	asn	val GTC	thr	16	val	AG.	r lys	55
] e	tyr lle TAT ATC	9. 6A	asp	918	£ 5	ž. ₹	¥	thr C ACC	a al
	1ys	tyr TAT	300 glu val glu asn asp glu met pro ala GAA GTG GAA AAT GAT GAG ATG CCT GCT	330 ala lys asp val phe leu gly met phe GCA AAG GAT GTC TTC TTG GCC ATG ITT	351 1ys thr tyr giu thr thr leu glu lys AAG ACA TAT GAA ACC ACT CTA GAG AAG	glu glu pro gln asn leu fle GAA GAG CCT CAG AAT TTA ATC	thr lys lys val pro gln val ACC AAG AAA GTA CCC CAA GTG	glu ala lys arg met pro GAA GCA AAA AGA ATG CCC	arg val thr lys AGA GTC ACC AAA	
_	. ser	1ys AAG		ala GCA	AC A	£ 3	thr.	_ ಕ್ರತ	2 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4	e e e e
35	231 val	261 ala GCC	291 ala GCC	321 91u GAG	351 175 AAG	381 val GTG	411 tyr TAC	441 070	471 asp GAC	501 91u 6A6

5	540 lys ala thr lys glu gln leu lys ala val met asp asp phe ala ala phe val glu lys cys cys lys AAG GCA ACA AAA GAG CAA CTG AAA GCT GTT ATG GAT GAT TTC GCT GCT TTT GTA GAG AAG TGC TGC AAG (1790)	570 phe ala glu glu gly lys lys leu val ala ala ser gln ala ala leu gly leu ter TTT GCC GAG GAG GGT AAA AAA CTT GTT GCT GCA AGT CAA GCT GCC TTA GGC TTA TAA CATCACATTTAAAAG (1883)	NAGAAAATGAAGATCAAAAGCTTATTCATCTGTTTTCTTTTCGTTGGTGTAAAGCCAACACCCTGTCTAAAAAGATAAATTTCTTTAA (2002)
10	ala ala phe val glu GCT GCT TTT GTA GAC	ala leu qiy leu ter GCC TTA GGC TTA TA	GCCAACACCCTGTCTAAAA
15	550 al met asp asp phe itt ATG CAT CAT TTC	580 ila ala ser gln ala kct GCA AGT CAA GCT	ICTTTTCGTTGGTGTAAA(
20	lu gin leu lys ala v AG CAA CTG AAA GCT C	ly lys lys leu val e GT AAA AAA CFT GTF (AAGCTTATTCATCTGTTTT
25	540 lys ala thr lys gi AAG GCA ACA AAA GA	570 phe ala glu glu g TTT GCC GAG GAG G	AAAGAAATGAAGATCAA
30	531 glu leu val lys his lys pro GAG CTC GTG AAA CAC AGG CCC	567 asp asp lys glu thr cys GAC GAT AAG GAG ACC TGC	ter ter CATCTCAGCCTACCATGAGAATAAGAGA
35	531 glu leu GAG CTC	561 ala asp GCT GAC	CATCTCAL

TCATTTTGCCTCTTTTCTCTGTGCTTCAATTAATAAAAATGGAAAGAATCTAA.... 20AA (2078)

Following are examples which illustrate procedures, including the best mode, for practicing the invention. These examples should not be construed as limiting. All percentages are by weight and all solvent mixture proportions are by volume unless otherwise noted.

Isolation of Messenger RNA Example 1

Human liver mRNA was obtained following the procedure of Chirgwin, et al [Chirgwin, J.M., Przybyla, A.E., MacDonald, R.J. and Rutter, W.J. (1979) Biochemistry 18, 5294-5299]. Immunoprecipitation of albumin containing polysomes was performed according to Taylor and Tse [Taylor, J.M. and Tse, T.P.H. (1976) J. Biol. Chem. 251, 7461-7467]. In vitro translation of mRNA was carried out in a reticulocyte cell-free system, following the instruction of the manufacturer (New England Nuclear). The translation products were separated electrophoretically according to Laemmli [Laemmli, J.K. (1970) Nature 227, 15 680-685.

Cloning Procedures Example 2

30

Double stranded cDNA was synthesized as described previously [Law, S., Tamaoki, T., Kreuzaler, F. and Dugaiczyk, A. (1980) Gene 10, 53-61]. It was annealed to PstI-linearized pBR322 DNA [Rolivar, F., 20 Rodriguez, R.L., Greene, P.J., Betlach, M.C., Heyneker, H.L., Boyer, H.W., Crossa, J.H. and Falkow, S. (1977) Gene 2, 95-113] that had been tailed with 15 dG residues/3'-terminus [Dugaiczyk, A., Robberson, D.L. and Ullrich, A. (1980) Biochemistry 19, 5869-5873]. The annealed DNA was used to transform E. coli strain RR1, as detailed previously [Law, S., et al., <a>Ibid.]. The albumin clones were selected using the colony hybridization method of Grunstein and Hogness [Grunstein, M. and Hogness, D.S. (1975) Proc. Natl. Acad. Sci. USA 72, 3961-3965], with [32p]-labeled cDNA synthesized with the immunoprecipitated polysomal mRNA as template.

As shown in Example 5, plasmids pHA36 and pHA206 were deposited in E. coli HB101 hosts. The plasmids were obtained from E. coli RR1 hosts, described in this example, and transformed into $\underline{\text{E.}}$ $\underline{\text{coli}}$ HR101 by standard procedures well known to those of ordinary skill in this The E. coli RR1 hosts were lysed and then centrifuged to 35 separate the chromosomal DNA, cell DNA and plasmid DNA. The plasmid DNA, remaining in the supernatant, is precipitated with ethanol and the precipitate is resuspended in buffer, e.g., TCM (10mM Tris·HCl, pH 8.0, 10 mM $CaCl_2$, 10 mM $MgCl_2$). The cells for transformation are

4083

prepared as follows: 120 ml of L-broth (1% tryptone, 0.5% yeast extract, 0.5% NaCl) are inoculated with an 18 hour culture of HB101 NRRL B-11371 and grown to an optical density of 0.6 at 600 nm. Cells are washed in cold 100 mM NaCl and resuspended for 15 minutes in 20 ml chilled 50 mM CaCl₂. Bacteria are then concentrated to one-tenth of this volume in CaCl₂ and mixed 2:1 (v:v) with annealed plasmid DNA, prepared as described above. After chilling the cell-DNA mixture for 15 minutes, it is heat shocked at 42°C for 2 minutes, then allowed to equilibrate at room temperature for ten minutes before addition of L-broth 10 times the volume of the cell-DNA suspension. Transformed cells are incubated in broth at 37°C for one hour before inoculating selective media (L-agar plus 10 µg/ml tetracycline) with 200 µl/plate. Plates are incubated at 37°C for 48 hours to allow the growth of transformants.

15 Example 3 Mapping of Restriction Endonuclease Sites

Restriction endonucleases were obtained from Bethesda Research Laboratories and New England Biolabs and were used according to the manufacturers' instructions. The digested DNA fragments were analyzed electrophoretically on agarose [Helling, R.B., Goodman, H.M. and Boyer, H.W. (1974) J. Virol. 14, 1235-1244] or acrylamide [Dingman, C., Fisher, M.P. and Kakefuda, T. (1972) Biochemistry 11, 1242-1250] gels.

Example 4 DNA Sequencing

phosphatase (Worthington) and labeled at the 5'-ends with polynucleotide kinase (Boehringer-Mannheim) and \(\gamma^{32p}\)ATP. Following digestion with a second restriction endonuclease and electrophoretic separation of the fragments, DNA sequence determination was done according to the procedure of Maxam and Gilbert [Maxam, A. and Gilbert, W. (1980) Methods Enzym. 65, 499-560] and the degradation products were separated electrophoretically on 0.4 mm acrylamide gels as described by Sanger and Coulson [Sanger, F. and Coulson, R. (1978) FEBS Letters 87, 107-110].

Example 5 Recombinant Plasmids pHA36 and pHA206

As disclosed in Example 2, albumin clones were selected by hybridizing to the enriched albumin cDNA probe. Plasmid pHA36 contained the largest insert of an albumin cDNA sequence. Both plasmids pHA36 and pHA206 have been deposited in a viable <u>E. coli</u> host in the

permanent collection of the Northern Regional Research Laboratory (NRRL), U.S. Department of Agriculture, Peoria, Illinois, U.S.A. Their accession numbers in this repository are as follows:

HB101(pHA36) - NRRL B-12551

5

HB101(pHA206) - NRRL B-12550

E. coli HB101 is a known and widely available host microbe. Its NRRL accession number is NRRL B-11371.

NRRL B-12550 and NRRL B-12551 are available to the public. upon the grant of a patent. It should be understood that the availability of these deposits does not constitute a license to practice the subject invention in derogation of patent rights granted with the subject instrument by governmental action.

E. coli RR1 and E. coli HB101 are known and widely available host microbes. Their NRRL accession numbers are NRRL B-12186 and NRRL
 B-11371, respectively.

pBR322 is a well known and widely available plasmid. It can be obtained from the following host deposit by standard procedures:

NRRL B-12014 - E. coli RR1 (pBR322).

YEp6 is a well known and widely available yeast episomal plasmid.

20 It can be obtained from the following host deposit by standard procedures:

E. coli HB101 (YEp6) - NRRL B-12093.

Example 6 Assembly of the Serum Albumin Gene

Assembling the pieces together is a straighforward task of restriction enzymology. There is only one MspI site in the overlapping
DNA sequence of the two cDNA clones. Two enzymatic steps of (i) MspI
digestion of the two DNAs, followed by (ii) the use of ligase, an
enzyme that seals DNA fragments, will give the desired product.
Although two other undesired DNA species will also be obtained in the
course of this recombination reaction, both of them will differ substantially in size. Thus, separation and isolation of the desired DNA
species will be achieved.

The assembled DNA clone can be used to transform two types of cells:

(a) Escherichia coli

- (b) Saccharomyces cerevisiae
- (a) The vector of choice is plasmid pBR322, the same that has

4083

been successfully used for cloning of the two fragmented pieces of the serum albumin cDNA.

(b) In order to transform yeast with the serum albumin structural gene sequence, the DNA must be inserted into one of the existing yeast plasmid vectors. This can be accomplished by taking advantage of the fact that several restriction endonuclease recognition sequences are absent from the cloned serum albumin DNA. Synthetic EcoRl DNA linkers can be ligated to the DNA fragment containing the serum albumin sequence followed by insertion (ligation) into one of the yeast plasmid vectors, e.g., YEp6, at the Eco Rl cloning site. The fused chimeric plasmid can be used to transform yeast according to an established procedure [Hinnen, A., Hicks, J.B. and Fink, G.R. (1978) Proc. Natl. Acad. Sci. USA, 75, 1929]. YEp6 can be obtained from the NRRL repository, as disclosed supra.

15 Example 7 Expression of the Serum Albumin Gene

The main body of the structural gene will be transcribed by the E. coli or yeast enzymes. If little or no albumin is produced with the selected host, then an Escherichia coli promoter DNA sequence carrying an initiation codon, i.e., ATG, can be ligated at the begin-20 ning of the serum albumin structural gene. Such elements are known and available, e.g., lac promoter used for the expression of human interferon gene in E. coli [Proc. Natl. Acad. Sci. 77, 5230 (1980)]; source of promoter DNA [Proc. Natl. Acad. Sci. 76, 760 (1979)]. Also, see Nature, Vol. 281, October 18, 1979. It has already been 25 documented that such Escherichia coli promoter sequences function well in the expression of foreign genes in Escherichia coli [Mercereau-Puijalon, O., Royal, A., Cami, B., Garapin, A., Krust, A., Gannon, I. and Kourilsky, P. (1978) Nature 275, 505; and Goeddel, D.V., Kleid, D.G., Bolivar, F., Heyneker, H.L., Yansura, D.G., Grea, R., Hirose, 30 T., Kraszewski, A., Itakura, K., and Riggs, A. (1979) Natl. Acad. Sci. USA 76, 106]. For expression in yeast, see Rose, M., Casadaban, M.J. and Botstein, D. (1981) Proc. Natl. Acad. Sci. USA 78, 2460 and 4466. Screening of Clones Producing Albumin Example 8

Immunological methods can be used to detect small amounts of albumin made in a bacterium. Flat disks of flexible polyvinyl are coated with the IgG fraction from an immune serum and the disks are pressed onto an agar plate so that antigen released from an <u>in situ</u> lysed microbial colony can bind to the fixed antibody. The plastic

disk is then incubated with the same total IgG fraction labeled with radioactive iodine so that other determinants on the bound antigen can in turn bind the iodinated antibody. Radioactive areas on the disk expose X-ray film during autoradiography and thus identify colonies producing the protein which is being screened for. Detailed protocols of this procedure have been published [Broome, S. and Gilbert, W. (1978) Proc. Natl. Acad. Sci. USA, 75, 2746]. The purification of human serum albumin can be accomplished by using procedures well known in the art. For example, procedures disclosed in a chapter by T. Peters: Purification and Properties of Serum Albumin, in: The Plasma Proteins, Putnam, Ed. Academic Press, New York, 1975, can be used.

The work described herein was all done in conformity with physical and biological containment requirements specified in the NIH Guidelines.

CLAIMS

- 1. Plasmid pHA36, having a restriction endonuclease pattern as shown in the drawing.
- 5
 2. Plasmid pHA206, having a restriction endonuclease pattern as shown in the drawing.
- 3. E. coli HB101 (pHA36) having the deposit accession number 10 NRRL B-12551.
 - 4. E. coli HB101 (pHA206) having the deposit accession number NRRL B-12550.
- 5. A microorganism modified to contain a nucleotide sequence coding for the amino acid sequence of human serum albumin; said nucleotide sequence is as follows:

20

25

	(30)	(170)	(192)	(350)	(049)	(330)	(620)	(710)	(300)
	AGC	20 179 AA	50 818 GCA	8 2 E	110 Dro	140 177	170 gln CAA	200 cys TGT	230 g lu GAA
	Dhe TTT	bhe TTC	phe TTT	thr	AAC	TTA TTA	169 Cys TGC		ه <mark>ا ه</mark>
5	1ec CTC 1	asn I	alu CAA	ala GCA	330	tyr TAC	168 cys TGT	leu lys CTC AAG	phe TTT
	phe 1	GAA J	thr ACT 0	Va.	GAT		£ \$		pro lys ala glu phe CCC AAA GCT GAG TTT
	1 E	2 ¥3	val GTA	thr	1ys AAA	1 X	A A	aln ara CAG AGA	ala CC1
	ę T	9 5 5 5	glu val GAA GTA			phe leu lys lys TTT TTG AAA AAA	phe		pro lys ala CCC AAA GCT
10		116 o	AAT	75 leu cys TTA TCC	ain his Caa cac	phe leu TTT TTG	ala GCT	ala lys GCC AAA	5 333
				1ys AAA	1eu		ala CCT	ser TCT	
		lys asb AAA GAT	leu val TTA GTG	asp 1	ohe TTC	glu thr	Lys AA	36T	arg AGA
		phe TT	173	7 53	101 105 105	£ 8	tyr IAT	878 GCT	olu CAG
		10 87.9 10.00	40 val GTA /	70 phe 111	100 101 qlu cys GAA TGC	130 asn AAT	160 arg tyr AGG TAT	190 173 AAG	220 leu ser gln CTG AGC CAG
15		his CAT	his CAT	7 T	AAT	ass GAC	1 ys	917	1 es
		ala (asp CAT	thr Acc	2 8	F.S.	ala lys CCT AAA	190 alu aly lys ala GAA GGG AAG GCT	220 arg leu ser gln arg bhe CGC CTG AGC CAG AGA TTT
	•		SA S		91 y S				ata CCT
		glu val GAG GTT	afe TT	leu his CTT CAT	pro CCT	ala phe GCT TTT	phe phe TTC TTT	glu leu arg asp GAA CTT CGG GAT	
20		ser AGT	e d	JCA TCA	glu pro GAA CCT	thr		le.	ala trp ala val CCA TGC CCA GTA
. 20		lys :	34 Cys TGT	NA AAA	CAA CAA	124 cys	glu leu leu GAA CTC CTT	58	t 15
		CAC	91,	asp lys GAC AAA		met ATG	5 A	asp	£ 55
		ala GCA	gln CAG	62 cys TGT	ala lys CCA AAA	val GTG	5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 7 5 7 5 7	leu CTC	1 × ×
		asp GAT	leu CTT		91 Cy3 TGT	asp	ala CC	1ys AAG	phe TTC
25	*	4. gra	30 tyr leu TAT CTT	60 glu asn GAA AAT	90 cys TGC	120 val GTT	150 tyr TAT	180 pro CCA	210 glu arg ela phe lys ala trp ala GAA AGA GCI TIC AAA GCA TGG GCA
				al a	as o	91c 6AG	phe TTT	leu TTG	glu arg GAA AGA
		phe arg	ala gin GCT CAG	ser ala TCA GCT	ala CCT	77 (CA		leu CTG	95 64
			phe TTT	g)u GAG	met	arg AGA	pro CCT	17 130 100	91.9 CGA
		gly GGT	ala GCC	asp	glu met GAA ATG	val GTG	arg his pro AGA CAT CCT	177 lys ala ala cys AAA GCT GCC TGC	phe 111
30		9-16 ACC	11e ATT	ala GCT	91y CGT	Jeu TTG	5 Y	ala GCT	lys AA
		- 50	15	val GTT	t yr 1 A T	క్టి స్ట	879 AGA	1ys AAA	fg CA
		-1 tyr ser TAT TCC	val	25 P	thr tyr ACC TAT	; pro arg leu val arg CCC CGA TTG GTG AGA	, als	asp GAT	ser leu gln lys phe gly AGT CTC CAA AAA TTT GGA
		ala GCT	leu 176	thr ACA	ags	leu CTC	11e	al a	
35	5	ser ala tyr a	21 ala GCC	17s AAA	81 819 CGT	asn AAC	18 18 18 18 18 18 18 18 18 18 18 18 18 1	171 a1a GCT	201 818 GCC

	(890)	(980)	(1070)	(1160)	(1250)	(1340)	(1430)	(1520)	(1610)	(1700)
5	260 1eu CTT	290 11e ATT	320 ala GCT	350 ala GCC	389 Leu CCT	410 879 CGT	440 h1s CAT	&70 Ser AGT	500 1ys	530 val GTT
	88P GAC	289 cys 160	tyr	lea CTT	970 CCT	val GTT	1ys AAA	val	PT0 CCC	
	arg ala AGG GCG	7 13 CAC	AAC	AGA AGA	1ys	leu TTA	438 cys TGT	818 CCA	val GTT	ala leu GCA CTT
:	875 AGG	367 TCC	1ys AAA	leu leu CTG CTG	alu phe GAA TTT	Jeu CTG	637 lys cys AAA TGT		ty TAC	thr
	335 CAC	1ys AA	316 0ys TGC	Jeu CTG	ale GAA	9] 9 8CG	1ys	lys thr AAA ACG	thr	£ 5
10	asp GAT	leu glu TTG GAA	asp val GAT GTT	ser val val leu leu leu ard TCT GTC GTG CTG CTG AGA	asp GAT	aln asn CAG AAT	430 ser arg asn leu qly lys val qly ser TCA AGA AAC CTA GGA AAA GTG GGC AGC	alu GAG	g) a	ain ile iys iys CAA ATC AAG AAA
	318 GCT	leu TTG		val GTG	phe TTC		91 y 660	h is	asp	1ys AAG
	253 cys TGT	pro leu CCT CTG	ser lys AGT AAG	val	lys val AAA GTG	lys phe AAA TTC	val GTG	Je 776	val GTC	11e ATC
	of P		ser	ser TCT	1ys AAA	1 ys	1ys AAA	val GTG	ale GAA	£ \$
15	250 253 leu leu glu cys CTG CTT GAA TGT	280 glu 1ys GAA AAA	glu GAA	tyr	370 tyr ala TAT GCC	\$00 glu tyr GAG TAC	41y GGA	460 461 leu oys val TTA TGT GTG	490 ala leu glu val GCT CTG GAA GTC	1 S
15			310 val GTT	340 asp GAT			430 1eu CTA		490 ala GCT	520 g1u GAG
	gly asp GGA GAT	278 279 cys cys TGC TGT	asp phe GAT TIT	his pro CAT CCT	369 qlu cys GAA TGC	ج کی ج	asn	gla	ser TCA	ser qlu lys TCT GAG AAG
					qlu	glu gln lev GAG CAG CTT	AGA	tyr leu ser val val leu asn TAT CTA TCC GTG GTC CTG AAC	cys phe IGC ITT	ser alu ICT GAG
	246 cys his TGC CAT	lys glu AAG GAA	ala ala GCT GCT	arg arg AGA AGG	pro his CCT CAT	g all	15 T	Jeu CTG	5. 15.	
20	2 2 4 6 0 4 3 1 GC	,			pro CCT	a lu GAG	glu val GAG GTC	val GTC	pro CCA	leu CTT
	245 glu cys GAA TGC	lys leu AAA CTG	ser leu TCA TTA	tyr ala TAT GCA	asp (CAT	phe TTT	910	val	55	514 asp ile cys thr leu GAT ATA TCC ACA CTT
	2 6 A		ser TCA	tyr TAT	ala GCA	glu leu GAG CTT	thr leu val ACT CTT GTA	ser 1 TCC	asn arg AAC AGG	514 11e cys ATA TGC
	his thr cac acc	Ser AGT	pro	916	ala CCT		CTT) leu		11e ATA
25		36T	o led	tyr	1 3 ala 7 ccc	392 1 Cys	thr ACT		val GTG	
25	240 lys val AAA GTC	r 11e	a asp r cAc	330 phe leu TTT TTG	361 9 cys 2 TGT	asn A AAT	Pro CCA	450 glu asp GAA GAC	480 ser leu val TCC TTG GTG	\$ 50 60 70 80 80 80 80 80 80 80 80 80 80 80 80 80
		270 p ser T TCG	300 0 ala T CCT		360 3 0ys 5 TGC	390 s gln	420 thr	450 910 1 GAA	480 3er	510 hts
	thr T Acc	n asp A GAT	r pro	ATG	1 lys	1 ys	ser 5 TCA	919 GCA	g lu	phe TTC
	b leu	₹3	glu met GAG ATG	2 6 C	leu glu CTA GAG	lev 11e TTA ATC	gln val CAA GTG	448 0ys	ACA SP	thr
30	ser lys leu val thr asp leu ICC ANG ITA GIG ACA GAT CTT	265 tyr lle cys glu asn gln TAT ATC TGT GAA AAT CAA	2 2	ala lys asp val phe leu gly GCA AAG GAT GTC TTC TTG GGC		gin asn leu lle CAG AAT TTA ATC	. 29 . 29	0 Pro	677 673 760	ala glu thr phe thr GCT GAA ACA TTC ACC
	G E	65	5 2	ag E	thr ACT	gin asn CAG AAT	. pro	met ATG	476 cys TGC	th Ag
	2 5	265 e cys	glu val glu asn asp .caa gtg gaa aat gat	. va	thr Acc	25	thr lys lys val ACC AAG AAA GTA	glu ela lÿs arg GAA GCA AAA AGA	thr 1ys ACC AAA	ala giu thr GCT GAA ACA
	1 5	7 7 7	₹§	3 455	916	pro	1 1ys	1.5 ¥	ACC CF	
	ÄŠ	3 tç	2 4	ala lys GCA AAG	thr tyr ACA TAT	g lu GAG	1ys	8 1 8 CCA	arg val AGA GTC	phe asn
35		261 ala lys GCC AAG	16 3.	1 ala	thr ACA	glu GAA				501 glu phe asn GAG TIT AAT
	231 val GTT	261 ala GCC	291 ala GCC	321 91u 6AG	351 1ys AAG	381 val CTC	411 tyr TAC	441 Pro	471 339 CAC	501 glu GAG

	(1790)	(63)	(2)
	Ξ	3.0	121
	559 560 cys 1ys TGC AAG	, WG	¥
5	559 673 167	tel	Ē
	558 cys 100	ACAI	IAATI
_	1ys AAG	CATC	Ž
	910	ter IX	AAA
	Val GTA	leo TTA	CTAA
10	a F	91.y	בדכד
	ala ncT	1 ×	Š
	ala GCT	313 CC	Ϋ́Υ
	T 6	ala GCT	AAAG
	540 Iys ala thr lys glu gln leu lys ala val met asp asp bhe ala ala phe val glu lys cys cys lys AMG GCA ACA AAA GAG CAA CTG AAA GCT GTT ATG GAT GAT TTC GCT GCT TTT GTA GAG AAG TGC TGC AAG	570 ala glu glu gly lys lys leu val ala ala ser gin ala ala leu gly leu ter GCC GAG GAG GGT AAA AAA CTT GTT GCT GCA AGT CAA GCT GCC TTA GGC TTA TAA CATCACATTTAAAAG (1883)	5757
15	asp GAT	AGT	5110
	ATG .	ala GCA	2110
	1 2 5	ala	101
	ata CCT	val CTT	111
20	<u>\$</u> }	Sec	57.5
	Tec	2.3	Z)L
	ale AA	2.3	TTA
	31c 34c	aly cor	X
	\$ \$	946 950	Y CA
25	S&O Chr ACA	570 glu GAG	¥
	* CA	2 U	MATC
	× 2	phe 71	Ş
	200	567 cys p TCC 1	§
	× × ×	h 3	ter
30	Ac /	% A	ter ter TGAGAATAA
	531 glu leu val lys his lys pro GAG CTC GTG AAA CAC AAG CCC	567 asp lys glu thr cys GAT AAG GAG ACC TGC	ter ter Catctcagcctagcatagagaaagaaagaatgaaggttattcatgtgtttttgtttttggtggtgaaggcaagaggtgtgtaaaaaagataaatttttaa (2002)
	al 1	sp 1 AT A	CTAC
	5 5 6	35 CAC 88	Š
	531 glu le GAG C	561 ala a: CCT C	ITCT
35	N 9 9	R B R	ð

TCATTTTGCCTCTTTTCTCTGTGCTTCAATTAATAAAAATGGAAAGAATCTAA..... 20AA (2078)

6. Nucleotide sequence of the cDNA of human serum albumin, said nucleotide sequence is as follows:

5	(170)	(192)	(350)	(440)	(330)	(620)	(710)	230 glu GAA (300)
	20 1ys	50 ala GCA	98 CTT	110 CCA	140 try TAT	17 17 18 18	200 cys TGT	230 glu GAA
		phe TTT	thr	AAC	3 4	169 160 160		
	asn bhe AAT TTC	3 5	ala GCA	88 54	TAC TAC	168 cys TGT	leu lys CTC AAG	phe ala TTT GCA
10		thr		asp	<u>\$</u> ₹	76. 64.	AGA AGA	ala alu phe GCT GAG TTT
10	glu glu GAA GAA		thr val ACA GTT	AAA	lys lys Aaa aaa	thr glu ACA GAA	aln ara CAG AGA	ala alu GCT GAG
		glu val GAA GTA		h I s	leu TTG			1 ys
	leu aly TTG GGA		75 leu cys TTA TGC	ain his CAA CAC	phe leu TTT TTG	ala phe GCT TTT	ala lys GCC AAA	pro lys CCC AAA
	asp leu aly GAT TTG GGA	val asn GTG AAT	1ys AAA	Jeu TTG	thr ACA	ala GCT	ser TCT	a per T T T T T T T T T T T T T T T T T T T
15	lys	40 glu asp his val lys leu val GAA GAT CAT GTA AAA TTA GTG	70 phe qly asp lys leu cys TTT CGA GAC AAA TTA TGC	phe leu TTC TTG	130 asn glu glu thr phe leu lys lys AAT GAA GAG ACA TTT TTG AAA AAA	lys ala ala phe AAA GCT GCT TTT	367 TCG	220 ser gin arg she AGC CAG AGA TTT
•	phe TTT	40 val lys leu GTA AAA TTA	91y CGA	101 cys TGC	alu GA	tyr TAT		glu
		40 val GTA	70 phe TTT	5 E 45	130 asn AAT	160 arg AGG	190 lys ala AAG GCT	220 Ser AGC
		his CAT	le CTT	88n AAT	asp GAC	1ys AAA	alu aly GAA GGG	
	glu val ala his GAG GTT GCT CAT	alu asp his GAA GAT CAT	thr	3	h is	ala CCT		trp ala val ala arg leu TGG GCA GTA GCT CGC CTG
20	val GTT	9 P	leu his CTT CAT	93	ala phe GCT TTT	phe phe TTC TTT	asp	ala GCT
	glu val GAG GTT	Phe TT		pro			819 000	val
	lys ser AAG AGT	676 CCA	36 TCA	g &	thr	leu leu CTC CTT	glu leu GAA CTT	trp ala TGC GCA
		94 633 1GT	lys AA	gla	124 cys TGC	leu CTC		tr 55
25	ale his GCA CAC	34 gln gln cys CAG CAG TGT	asp GAC	ala lys GCA AAA	val met GTG ATG	ole GAA	asp	ala CCA
			62 0y3 1GT			970 CCG	leu CTC	1ys AAA
	asp GAT	30 tyr leu TAT CTT	60 glu asn GAA AAT	90 91 cys ays TGC TGT	asp	818 CCC	1ys AAG	phe TTC
		30 tyr TAT			120 val GTT	150 tyr 1A1	180 CCA	210 818 GCT
		gln	ala GCT	asp GAC	pro glu	phe TTT	1eu 77G	arg AGA
30		ala GCT	164 164	ala GCT	pro CCA	tyr TAC	leu CTG	3
		phe TTT	g GAG	met ATG	ACS A	pro CCT	5. 13. 13.	aly SGA
•		a1a GCC	asp GAT	glu met GAA ATG	leu val TTG GTG	his	, e 5	phe 111
		11e ATT	ala GCT	gly cct	Jeu 116	arg AGA	ala GCT	15.8 AA
25		leu TTG	val GTT	tyr TAT	5 45 75	arg AGA	1 ys	g V
35		val	53 cys val TGT GTT	thr	014 CCC	818 CCC	asp	Jeiu CTC
		Jeu TTG	thr	9}n	leu CTC	11e ATT	ala GCT	Ser
		21 ala GCC	51 175 AAA	81 arg	111 asn AAC	돌물물	171 ala GCT	201 ala CCC

5	(890)	(980)	(1070)	(1160)	(1250)	(1340)	(1430)	(1520)	(1610)	(1700)
	260 leu CTT	290 11e ATT	320 ala CCT	350 ala GCC	380 leu CCT	4t0 ara ccr	880 818 CAT	470 3er Agt	Sno 1ys AAA	530 val GTT (
	435 GAC	289 cys TCC	tyr IAT	3 E	51 0	val	£ \$	val :	010	CTT C
	818 606	11.8 CAC	asn AAC	ដ្ឋ ស្	lys AAA	7 E	438 cys	ore CCA (val p	ala leu GCA CTT
10	5. 53 5. 53	36. 19.	1ys AAA	. p. c.	phe 111	Jeu	437 cys TGT	thr	tyr Y	thr ala ACT GCA
	ass GAC	1ys AAA	316 cys TGC	leu leu CTG CTG	o jo CAA	8 L 8				ain thr CAA ACT
	asp GAT	مارة. 1979	val GTT	leu CTG	335	asn	ser lys AGC AAA	alu lys CAG AAA	alu thr GAA ACA	lys c
	ala CCT	leu leu CTG TTG	asp	val GTG		16. CA5			asb GAT	Iys Iys AAG AAA
	253 cys TGT	Jeu CTG	lys AAG	val GTC	val	phe TTC	val	TG 1	val a	1
15	leu glu CTT GAA	pro	Ser	3er 101	173 AA	1 ys	1ys	val	glu val GAA GTC	ain ile CAA ATC
	250 leu leu CTG CTT	280 qlu lys GAA AAA	ale GAA	tyr TAC	a) a	tyr 7AC	' 5		Jeu CTG	
	250 leu CTG		350 val	340 asp cat	370 tyr TAT	400 91u GAG	430 1eu CTA	460 461 leu cys TTA TGT	h90 ala leu GCT CTG	520 glu arq GAG AGA
	asp CAT	279 cys TGT	phe TTT	Pro CCT	369 cys 1GC) te:	asn AAC		Ser TCA	
20	91 y 664	278 cys 1GC	asp GAT	N IS	his alu CAT GAA	leu CTT	arg AGA	ser val val leu asn gln TCC GTG GTC CTG AAC CAG	phe TTT	glu lys CAG AAG
	S his	910	ala . ccT	arg arg AGA AGG	h is	aln CAG	36. TCA	leu CTG	cys 1GC	ser TCT
	5 246 cys : TGC	leu lys CTG AAG	ala GCT	arg AGA	pro CCT	glu	glu val	val GTC	pro	leu CTT
	245 1 cys 1 TCC		3 7	# [# CCA	asp	p ge		val GTG	8.79 CGA	514 cys thr TCC ACA
	ole c	ser lys AGT AAA	ser TCA	tyr TAT	ala GCA	leu CTT	val GTA	Ser TCC	arg	514 cys TGC
25	thr ACG		pro	ole CA	ala GCT	9. CAS	thr leu val ACT CTT GTA	ST V	asn AAC	11e ATA
	val his GTC CAC	3er TCC	leu 11G	tyr TAT	8 8 000	392 cys TGT	thr	1 XT		asp ile GAT ATA
			359	1eu	361 cys TGT	AAT	5 CCA	asa GAC	leu val TTG GTG	ala GCA
	240 1ys	270 3er	300 ala CCT	330 phe TTT	360 cys 160	330 91n CA	420 thr ACT	450 glu GAA	\$6r 7CC	510 his ala cat gca
20	thr Acc	1 85P	pro CCT	met ATG	1ys AAG	lys AA	ser TCA	8 te 60A	of g	phe TTC
30	2 L	asn gin AAT CAA	glu met CAG ATG	leu gly TTG GGC	leu glu CTA GAG	leu ile TTA ATC	va1 GTG	448 033 1GT		
	ser lys leu val thr asp leu TCC AAG TTA GTG ACA GAT CTT	glu asn GAA AAT	916		leu CTA	leu TTA	gla	era CCC	477 cys TGC	glu thr phe thr GAA ACA TTC ACC
	1 th		asp CAT	phe TTC	thr ACT	gin asn CAG AAT	. F 000	glu ala lys arg met pro GAA GCA AAA AGA ATG CCC	476 Cys TGC	glu thr GAA ACA
	\$ N	265 cy3 TGT	glu asn GAA AAT		thr	gIn CAG	val	arg AGA	₹ ¥	
35	2 E	tyr 11e TAT ATC		iys asp AAG GAT	glu GA	glu pro CAG CCT	17s	<u>₹</u> ¥	thr	ala CCT
	¥	tyr TAT	val GTG	1ys AAG	tyr IAT	głu GAG	lys AAG	الا 20 م	val	asn AAT
	1 1 Se1	1 1ys	g to CAA	ala GCA	351 lys thr AAG ACA	91u CAA	411 tyr thr lys lys val pro gln val TAC ACC AAG AAA GTA CCC CAA GTG	914	arg val thr lys AGA GTC ACC AAA	501 glu phe asn ala GAG TTT AAT GCT
	231 val GTT	261 ala GCC	291 ala GCC	321 glu GAG	351 lys AAG	381 val GTG	411 tyr 1AC	441 pro	asp GAC	501 glu GAG

5	(1790)	(1883)	ter ter ter Cateleagaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaa	
	560 178	¥c	¥	
	559 978 1GC	ter TAA	TCTT	
	558 cys TGC	ACAT	AATT	
	Y S	CATC	CATA	
10	36.1	T &	AAAA	
	781 GTA	1 teu	CTAA	
	a E	50	CTCT	
	ele r	1 E	וכעככ	
15	ele CCT	a1a 505	γυ	
	Phe TTC	ala GCT	LAAA	
	350 350 GAT	580 gln CAA	cctc	8
	550 giu gin leu lys ala val met asp asp phe ala ala phe val giu lys cys dys l . GAG CAA CTG AAA GCT GTT ATG GAT GAT TTC GCT RTT GTA GAG AAG TGC TGC /	570 glu glu qly lys lys leu val ala ala ser gln ala ala leu qly leu ter GAG GAG GGT AAA AAA CTT GTT GCT GCA AGT CAA GCT GCC TTA GGC TTA TAA CATCACATTTAAAAG	CGTT	TAATAAAAATGGAAAGAATCTAA 20AA (2078)
	#et ATG	e le CCA	1111	¥
20	val GTT	8 s c c c c c c c c c c c c c c c c c c	11101	:
	als CCT	val GTT	IGTT	8
	AX AX	CTI CTI	CATC	
	1 Le	8 1 ye	TATT	CTAA
25	£ 3	× 1×	AGCT	₹
	92	45 2	Υ Υ	¥¥
	540 ala thr lys c	6 g	AGAT	AATG
	Sto thr	570 a 91u c GAG	ATGA	AAAA
		e ala T CCC	GAAA	TAAT
30		7 is 5.	8	[CAA]
	# C	567 r cys IC TGC	AQ.	70
	% T	E A	Ç.	2
	∓ 5	6 9 6 9	ATG.	
35	leu val lys his lys pro lys CTC GTG AAA CAC AAG CCC AAG	567 e asp lys glu thr cys phe ic gar AAG GAG ACC TGC TTT (CTAC(TARULTACTURE
	בן ק בן	asp as GAC G/	CAGG	
	531 glu le GAG CT	561 ala a: GCT &	ATCT	1
	N 29	Ø 8 €	ű	1

TCATTTTGCCTCTTTTCTCTGTGCTTCAATTAAAAAATGGAAAGAATCTAA.... 20AA (2078

7. Nucleotide sequence coding for the prepeptide of human serum albumin, said nucleotide sequence is as follows:

	albumin,	said	nucleotide	sequence
				(30)
5				AGC
				ohe TTT
				-10 leu leu phe leu CTT CTT TTT CTC
				phe
10				O CTT
				leu l
				ser TCC
				t o tlu phe 11e s ACC TTT ATT 1
				ag E
15				r o 1 tlu A ACC
				p r trp val
				6 s 7 t
				3 t 3 ≥ 3.
20				1 2 2
				### ##################################
				מכנכ
				CACA
25				YCC
				rerev
				נדזכו
				1101
30				6CTT.
30				_

ser ale tyr ser arg gly val phe arg ard TCG GCT TAT TCC AGG GGT GTG TTT CGT CGA

35

8. Nucleotide sequence coding for pro human serum albumin, said nucleotide sequence is as follows:

5	(170)	(1560)	(350)	(999)	(330)	(620)	(017)	(300)
	20 173 AAA (50 ala CA 7	80 1eu CTT (110 pro ccA (140 try TAT	170 GA C	200 cys 7GT	230 glu GAA
	ohe 1 TTC /	phe a	thr 1	AAC	15 T	169 cys TGC	1ys	
	AAT	g Iu g	818 GCA	330	tyr TAC	168 cys TGT	leu CTC	
	9 to 0	thr o	715 CTT	GAT	17s	ale GA		glu
10	alv alv GAA GAA	val	thr	1ys AAA	1ys	AC th	ain ara CAG AGA	ala qlu phe GCT GAG TTT
	41. 66.	glu val GAA GTA			1 tc	. ag 1	1ys	
	leu aly TTG GGA	asn glu AAT GAA	75 asp lys leu ays GAC AAA TTA TGC	leu ain his TTG CAA CAC	a t L	ala GCT	ala CCC	סדס גוננ
	885 GAT		1ys	leu TTG	ACA ACA	ala GCT	ser TCT	ohe TT
15	10 arg phe lys asp CGG TTT AAA GAT	leu val TTA GTG	asp GAC	ohe TC	5 5	1ys AAA	36T TCG	arg AGA
•	phe	1 ys AAA	g g CGA	101 8,53 16,5	g Ju GAA	tyr	ala GCT	gln
	10 20 20 20 20	40 val lys GTA AAA	70 phe gly TTT CCA	100 GA 6	130 asn AAT	160 819 AGG	190 173 AG	220 Ser AGC
		his CAT	1 ec		asp GAC	lys AAA	9 to 200	leu CTG
	ala his GCT CAT	asp GAT	thr	arg asn AGA AAT	h i s CAT	ala GCT	ale GA	arg CCC
20	val GTT	ole GAA		41p	phe TTT	ahe TT	88p GAT	ala GCT
	g) c	phe TT	leu his CTT CAT	pro	ala GCT	pro glu leu leu phé CCG GAA CTC CTT TTC	500 500	val GTA
	AGT	pro	ser TCA	g)u GAA	thr	leu CTT	glu leu GAA CTT	ala GCA
	1ys AAG	34 cys TGT	lys AAA	gla	124 cys 760	CTC		trp 166
25	M S CAC	glı	asp GAC	1ys	met ATG	pro glu CCG GAA	asp CAT	ala GCA
25	* 1 ° C	gla	62 0ys 1GT	£ 55	val GTG		leu CTC	lys AAA
	asp GAT	leu CTT	asn	93 167 167	asp GAT	818 000	1ys AAG	phe TTC
	- 500	30 tyr TAT	60 916 GAA	90 93 760	120 val GTT	150 tyr 1A1	180 Pro CCA	210 ala GCT
	o phe arg TTT CGT	gln CAG	ala GCT	asp GAC	alu GAG	tyr phe TAC TTT	leu TTG	arg AGA
30	phe TTT	ala GCT	36T	ala GCT	pro CCA		leu CTG	gle GA
	p	phe TTT	glu	met ATG	arg AGA	pro CCT	55 160	gly GGA
	gly GGT	ala CCC	asp	gly glu met GGT GAA ATG	teu val arg TTG GTG AGA	arg his pro AGA CAT CCT	177 ala ala cys GCT GCC TGC	leu gln 1ys phe CTC CAA AAA TTT
	9-19 AGG A	11e ATT	al a	aly cct	14 TG	arg AGA	ala GCT	173 AA
		leu TTG	val GTT	tyr	క్ట్ ప్ర	arg AGA	<u>₹</u>	ala CAA
35		21 ala leu val leu lle ala GCC TTG GTG TTG ATT GCC	53 9ys TGT	thr tyr	pro arg	ala arg GCC AGA		leu CTC
		leu TTG	thr	£6 ₹ 9	Jeu CTC	141 glu 11e GAA ATT	ala GCT	Ser
		21 818 500	51 AA	81 879 CCT	111 asn AAC	141 910 CAA	171 a1a GCT	201 ala GCC

5	(890)	(980)	(1070)	(1160)	(1250)	(1340)	(1430)	(1520)	(1610)	(1700)
	260 1eu CTT	290 11e ATT	320 ala CCT	350 ala GCC	380 160 CCT	410 ara cGT	A40 hIs CAT	470 ser AGT	500 1ys AAA (530 val GTT (
,	ala asp ccc cAc	283 157 157	tyr TAT	arg leu AGA CTT	pro			V81	bro CCC	lea C
		h is	ASO AAC		1ys AA	Jeu TTA	438 cys TGT	878 CCA	val	ala 1 GCA 0
10	asp arg GAC AGG	15 T	316 cys lys tgc aaa	leu CTG	phe TTT	leu leu val CTG TTA GTT	437 635 161	thr	tyr TAC (thr a
	ala asp asp arg CCT GAT GAC AGG	1ys		leu leu leu CTG CTG CTG	26. AA	ala ac			thr	lys gin the AAA CAA ACT
	5 7	ala GA	val GT	leu CTG	335 GAT	asn AAT	ser AGC	alu lys GAG AAA	alu thr GAA ACA	lys g
		Jeu 116	asp GAT	val	phe TTC	phe ain asn ala TTC CAG AAT CCG	4 1 y	F S		
	253 Cys	leu CTG	ser lys asp val AGT AAC GAT GTT	val	val	age 7	val GTG	1ec	val GTC	11e ATC
15	28	lys pro leu AAA CCT CTG	Ser	ser TCT	ala lys GCC AAA	173	1ys AAA	val GTG	g g	ain lie lys CAA ATC AAG
	250 leu leu CTG CTT		310 val glu GTT GAA	tyr TAC		400 glu tyr GAG TAC		461 Cys TGT	490 ala leu glu val asp GCT CTG GAA GTC GAT	75
	250 1 eu CTG	280 910 6AA	310 val	340 asp GAT	370 tyr TAT	400 91c GAG	436 asn leu AAC CTA	\$60 1eu 17A	ala GCT	520 91u GAG
	83p	278 279 0ya cya TCC TCT	phe TTT	pro CCT	369 cys 1GC	9. Y S	AAC	gla	ser TCA	1ys AAG
20	s 91y r GGA	278 J 093 A TGC	3 85p	arg his AGG CAT	alu GAA	gln leu CAG CTT	ser ang TCA AGA	asn AAC	phe	910
	6 B hts C CAT	9 9 CA	s ala		his CAT			Jeu CTG	cys 700	ser TCT
	5 246 8 cys 5 TGC	J lys	ala CCT	9 3 C	P70	alu GAG	glu val GAG GTC	va1 070	5 Y	leu CTT
	245 U cy8 A TGC	a leu A CTG	TA TTA	818 -	asp CAT	phe TTT		vs1 GTG	ఫ్ల్ సై	th ACA
	5 9 5 A 5 A 5 A 5 A 5 A 5 A 5 A 5 A 5 A	ser lys	s ser	glu tyr GAA TAT	8 g s	glu leu GAG CTT	leu val CTT GTA	3er 700	asn arg arg AAC AGG CGA	514 cys TGC
25	240 lys val his thr AAA GTC CAC ACG	r ser C AGT	u pro		ala GCT			leu CTA	asn	11e ATA
	i his c cac	e Ber	p leu C TTG	tyr 174	ala OCC	392 cys	thr ACT	tyr TAT	val GTG	asp GAT
.	240 lys val AAA GTC	r 11e G ATC	a asp T CAC	1 1 1 C	361 cys : TGT	asn AAT	D D CCA	889 GAC	480 ser leu TCC TTG	ala CCA
	240 r 1ys c AAA	270 p ser T TCG	300 8 8 8 8 7 CCT	330 t phe	360 933 1GC	390 1 gln	420 thr	450 91u GAA	480 ser TCC	510 hfs CAT
20	u thr T ACC	n asp A GAT	t pro	y met	1 1ys	¥	ser TCA	8 1 8 GCA	g Ju GAA	phe TTC
30	p leu T CTT	n gln T CAA	glu met CAC ATG	phe leu gly TTC TTG GGC	leu glu CTA GAG	· Te	pro gln val CCC CAA GTG	448 cys TGT	477 cys thr TGC ACA	th Acc
	thr asp ACA GAT	glu asn GAA AAT	68	e Jeu	thr leu ACT CTA	1 Fe	. g1	979 CCC	477 cys TGC	phe TTC
•		s glu r GA	1 83p			88 AA1		met ATG	476 Cys TCC	glu thr GAA ACA
	8 Z	265 11e oys ATC TGT	glu asn CAA AAT	asp val GAT GTC	glu thr	4 2	val GTA	lys arg AAA AGA	173 AA	glu
35	ser lys leu val TCC AAG TTA GTG		£ 3		95 84	glu glu pro gln asn leu lle GAA GAG CCT CAG AAT TTA ATC	411 tyr thr lys lys val TAC ACC AAG AAA GTA	lys AAA	thr	ala GCT
	Ž.Ž	lys tyr AAG TAT	glu val GAA GTG	ala lys GCA AAG	tyr TAT	glu	Lys	55 50 50 50 50 50 50 50 50 50 50 50 50 5	val	
		1 8 1ys C AAG	a glu c GAA	ata ccA	351 1ys thr AAG ACA	gra	thr	9 to 25	arg AGA	phe TTT
	231 val GTT	261 818 GCC	291 ala GCC	321 91u GAG	351 173 AAG	381 val GTG	411 tyr TAC	441 910 CCT	471 889 GAC	501 glu GAG

5	(1790)	570 ala glu glu qly lys lys leu val ala ala ser gln ala ala leu qly leu ter GCC GAG GAG GGT AAA AAA CTT GTT GCT GCA AGT CAA GCT GCC TTA GGC TTA TAA CATCACATTTAAAAG (1883)	ter ter Catctcagcctaccatgagaataagagaagaaaaaaagatcaaaagcttattcatctgtttttctttttcgttggtgaaagccaacacctgtctaaaaaacataaatttctttaa	
	350 173 4AG	\$€	TA.	
	558 559 560 oys cys lys TGC TGC AAG	ter ITA	TCT	
	558 575 575	ACAŢ	AATT	
	× × ×	SATC	CATA	
10) 1 J	± ₹	AAA	
•	181 g	3 5	CTAA	
	4 F	<u> </u>	CTGT(
	la p	3 ₹) YCC	
	18 g		CAAC	
15	ت ۾ پ	la a	AAGC	
	550 asp pl	580 g1n a CAA G	rctA	
	A 45 G 9. 55	61 5 C C G	TTC0	(8/0;
	540 ala thr lys glu gin leu lys ala val met asp asp phe ala ala phe val glu lys cys cys lys GCA ACA AAA GAG CAA CTG AAA GCT GTT ATG GAT GAT TTC GCT CCT TTT GTA GAG AAG TGC TGC AAG	s ₹	TTC	TTATTAAAAATGGAAAGAATCTAA 20AA (2078)
20	E Y	ا الا ع	CTTT	•
	2	18 18 18 18 18 18 18 18 18 18 18 18 18 1	E	0
	. o %.₹	3 =	cr61	:
	20 7 9	~ \$ %	TCA1	×
	. Y	2 4	TTA	\TCT/
25	AG G	7 5 7 4	AAGC	ΑGA
	£.≷). AG G	TCAA	₹
	5 ± 5	570 ala glu glu GCC GAG GAG	¥ĕ	AAT
	¥ ¢ % ≿ <u>®</u>	ν υ υ	AATG	¥
	و م الا	G a ⊥ e	¥ç	TTAA
30	6 H	567 cys pl TGC T	¥ĕ	TCA
	\$ 9 C 5	7 2 2 E	ter TAAG	TCCT
	* O	5 E	Ş	Tric
	val 1ys his 1ys pro 1ys GGG AAA CAC AAG CCC AAG	رة ع و ع	te SATO	
35	. ~ \$	5 N	TAC	7
3 3	G %	5 5 5 8 5 8	אפכנ	į
	531 glu leu v GAG CTC (561 ala asp asp 1ys glu thr cys phe GCT GAC GAT AAG GAG ACC TGC TTT	VTCT	
	8 2 2	561 e16 CCT	ວັ	ì

1CATTTTGCCTCTTTTCTCTGTGCTTCAATTAATAAAAATGG

9. Nucleotide sequence coding for the pre pro human serum albumin, said nucleotide sequence is as follows:

5	(30)	(170)	(1260)	(350)	(660)	(330)	(420)	(710)	(300)
	AGC	20 1ys AAA (50 % SO	80 1eu CTT	110 670 CCA	160 try TAT	170 91n CAA	200 cys TCT	230 glu GAA (
	a E			thr 1	AAC		169 cys 100		
	1eu a	asn phe AAT TTC	alu phe CAA TTT	ala t GCA A	8 0 Y	tyr leu TAC TTA	168 cys 101	leu lys CTC AAG	phe ala TTT GCA
	phe 1	g ble a	thr a		aso a	, ys (glu GAA		alu n
10	3 5	g	al t	thr v		1 × 1	th And	aln CAG J	ala GCT C
	-10 lev lev CTT CTT	÷ 5	plu val CAA GTA	× 25 5	his lys Cac aaa	thr bhe leu lys lys ACA TTT TTG AAA AAA	phe thr alu TTT ACA GAA	ser ala lys aln ara TCT GCC AAA CAG AGA	phe pro lys ala glu phe ala TTT CCC AAA GCT GAG TTT GCA
	100	asp leu aly GAT TTG GGA	AAT	75 leu cys TTA TCC	gin his CAA CAC	a E	ala GCT	2 22	200
	lle ser Att TCC	SP 1	16 J	lys I	leu aln TTG CAA	4 5	ala d	ser ala ICT GCC	Dhe T
15			40 glu asp his val lys leu val asn GAA GAT CAT GTA AAA TTA GTG AAT	75 asp lys leu cys thr val GAC AAA TTA TGC ACA GFT		130 asn qlu qlu thr AAT GAA GAG ACA	160 arg tyr lys ala AGG TAT AAA GCT	36T	ACA PTG
	p r c lys trp val tlu phe AAG TGG GTA ACC TTT	10 arg phe lys CGG TTT AAA	AA I	÷ 5		130 asn qlu qlu AAT GAA GAG	160 arg tyr lys AGG TAT AAA	190 lys ala ser AAG GCT TCG	220 ser gin arg AGC CAG AGA
	2 5 5 7 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5	5 # 8 8 8	val GTA	70 phe qly TTT CCA	100 101 alu eys CAA TCC	130 asn AAT	160 ACC	190 lys ala AAG GCT	220 AGC
	tra 766	M.S.	MES CAT	1 E	AAT	asb GAC		91. 666	leu CTG
	1ye	818 GCT	asp CAT	thr Acc	2 3	h is CAT	ala lys GCT AAA	글	61e
20	Het ATG	vel ele GTT GCT	GA A	his CAT	9 9 555	phe		esp GAT	ala trp ala val ala arg leu GCA TGG GCA GTA GCT CGC CTG
		glu vel GAG GTT	phe TTT	lys ser leu his AAA TCA CTT CAT	pro	ala GCT	phe TTC	glu leu arg asp GAA CTT CGG GAT	val ala GTA GCT
		AGT	2 Y	3er 7CA	SA S	thr	le CTT	leu CTT	trp ala TGG GCA
		1ys	34 cy3 TGT	asp lys GAC AAA	g],	124 cys TGC	3 5	9f6 G&A	trp 166
25		ele hís GCA CAC	gln CAG	62 cys asp TGT GAC	1 ys	met ATG	glu GAA	asp GAT	lys ala AAA GCA
		ele hís GCA CAC	gla	62 cys	* 55	val	ala pro OCC CCG	leu CTC	lys AA
		- d d S	leu CTT	asn AAT	90 91 cys cys 818 TGC TGT GCA	asp	# T# CCC	1ys AG	phe TTC
		- £3	30 tyr 1AT	8 2 3	8 \$ 5	120 val GTT	150 tyr TAT	180 CCA	210 ala GCT
		p r o gly val phe arg GGT GTG TTT CGT	gh	ser ala g TCA GCT O	ala asp CCT GAC	g r GAG	phe TTT	leu leu CTG TTG	glu arg
30		1 & °	ala GCT	ser TCA	313 GCT	979 CCA	tyr TAC	Jeu CTG	9 8 84
		7	phe TT	5 2	glu met GAA ATG	arg AGA	pro CCT		91. 89.
		aly CCT	ala cc	asp GAT	£ 3	leu val TTG GTG	arg his AGA CAT	818 CC	phe
		9 5 9 V 2 9	11e ATT	ala GCT	31,4		8 3 8	ala GCT	1ys AAA
25		- <u>. 5</u> 5	35	val CTT	tyr TAT	37. (CA	87.9 AC	173 AA	gla
35		aer ala tyr ser	leu val leu lle TTG GTG TTG ATT	£ 57 € 75 € 75 € 75 € 75 € 75 € 75 € 75	glu thr tyr c	leu pro arg CTC CCC CGA	ala arg GCC AGA	esp lys	leu gin lys r
		ele GCT	15c	51 lys thr AAA ACA	26 ₹3	Je CTC	11e ATT	ala GCT	Ser
		367	21 al a	22 Lys	arg CGT	111 asn AAC	19 to 18 to	171 818 CCT	201 818 GCC

5	(890)	(086)	320 ala cr.t (1070)	(1160)	(1250)	(1340)	(1430)	(1520)	(1610)	(1700)
	260 1eu CTT	290 11e ATT	320 ala CCT	350 ala GCC	380 Teu CCT	410 ara CGT	860 613 CAT	470 Ser AGT	500 175 AAA	val GTT
	SAC CAC	289 cys 100	tyr	15	oro CCT		1ys AAA	ct A	679	Jes CTT
	- 9 - 9 - 8		AAC	25		1 te	438 cys TGT	870 CCA	val	ala GCA
	arg ala Agg OCG	ser his TCC CAC	lys asn AAA AAC	leu ard CTG AGA	phe 1ys TTT AAA	Jeu	437 cys TGT	thr ACG	17 17C	thr
10	SE SE		316 cys lys TGC AAA	Jeu CTG	a lu CAA	8 J 8	1ys AAA	glu lys GAG AAA	thr	£ 83
	asp asp GAT GAC	glu lys GAA AAA	316 val cys GTT TGC	Jeu CTG	GAT	asn ala AAT CCG	AGC	alu GAG	est of the second	lys lys AAG AAA
	8 8 8	15 T	asp val	val	ohe TTC	aln CAG	چ وو ځ	h1s CAT	asp	1ys AAG
	233 673 767	pro leu leu glu CCT CTG TTG GAA	lys asp AAG GAT	ser val val leu leu leu ard TCT GTC GTG CTG CTG AGA		ghe TTC	val GTG	Jec 77G	val GTC	520 glu lys glu arq qin ile lys lys qin thr ala leu GAG AAG GAG AGA CAA ATC AAG AAA CAA ACT GCA CTT
15	SA SA	970 CCT	Ser	Ser	lys val AAA GTG	lys AA	1ys AAA	val GTG	916	ala CA
	leu CTT	1 ys	glu ser GAA AGT	1 yr	370 tyr ala lys TAT GCC AAA	400 glu tyr GAG TAC	630 ser arg asn leu qly lys val TCA AGA AAC CTA GGA AAA GTG	460 461 Jeu cys TTA TGT	490 ala leu CCT CTG	520 glu lys glu arg GAG AAG GAG AGA
	250 1eu CTG	S = 28	310 val GTT	340 asp GAT	370 tyr ala TAT GCC	91c 6Ac	430 1eu CTA	460 461 Jeu cys TTA TGT	ala CCT	520 glu GAG
	245 246 250 oys cys cys his gly asp leu TGC CAT GGA GAT CTG	279 cys TGT	310 phe vel TTT GTT	pro CCT	369 cys TGC	91 × 00 × 00 × 00 × 00 × 00 × 00 × 00 ×	AAC	tyr leu ser val val leu asn gln TAT CTA TCC GTG GTC CTG AAC CAG	phe ser TTT TCA	1ys AAG
	91 66A	278 575 100	asp GAT	h is	a) a	glu gin leu gly GAG CAG CTT GGA	AGA A	ASO	cys phe ICC TIT	910
20	his gly CAT GGA	₹.8	ala ac T	arg AGG	his CAT	alu aln GAG CAG	ser	Jeu CTG	698 166	ser TCT
	246 973 760	lys AAG	8 3 CCT	arg AGA	pro CCT	a)u GAG	leu val glu val CTT GTA GAG GTC	5.81	675 CC A	514 gys thr leu ser TCC ACA CTT TCT
	245 978 160	ser lys leu lys AGT AAA CTG AAG	leu TTA	ala GCA	asp CAT	phe TTT	976	val GTG	879 693	thr
	3	1ys	36. 1CA	glu tyr GAA TAT	چ در کی	Jeu CTT	val GTA	ser TCC	arg AGG	514 6ys 100
25	thr Acc	ser lys AGT AAA	Pro CCT		ala CCT	glu leu GAG CTT	leu CTT	Jec CTA	asn	asp ile CAT ATA
	÷ 5	ser TCC	leu TTG	tyr TAT	al a	392 cys 1GT	thr	tyr TAT	val GTG	asp GAT
		11e ATC	CAC	leu TTG	360 361 0ys cys TGC TGT	390 gIn asn CAA AAT	pro	450 glu asp GAA GAC	1et	510 his ale cat gca
	240 1ye vel AAA GTC	270 3er 106	300 ala GCT	330 phe TTT	360 361 1ys oys cys AAG TGC TGT		420 thr ACT	450 g1u GAA	480 ser TCC	510 h1s CAT
	thr	asp	pro	met ATG		¥	ser TCA	ele CCA	glu	phe
30 35		g C¥3		91y GGC	glu GAG	Te ATC	gln val CAA GTG	448 cys TGT	thr	thr
		asn	glu met GAG ATG	val phe leu gly GTC TTC TTG GGC	glu thr thr leu glu GAA ACC ACT CTA GAG	leu TT	91. CA	010	477 cys TGC	phe TTC
	thr ACA	giu asn GAA AAT	asp CAT	phe TTC	thr	asn AAT	vel pro	met ATG	476 cys TGC	thr
	leu val thr TTA GTG ACA	265 cys TGT	asn AT	val GTC	YCC Y	g CAG	V. STA	arg AGA	val thr lys GTC ACC AAA	_ g &
	1ec 11A	11e	g GA	asp GAT		pro CCT	lys lys AAG AAA	17s	th Acc	ala GCT
	¥. 84. 84. 84. 84. 84. 84. 84. 84. 84. 84	tyr	val GTG	ala lys GCA AAG	174	glu glu pro gin asn leu ile GAA GAG CCT CAG AAT TTA ATC	1ys AG	glu ala Iys arg met GAA GCA AAA AGA ATG	val GTC	asn AAT
	ser lys leu val thr asp TCC AAG TTA GTG ACA GAT	lys AG	glu GAA	ala GCA	thr	ole CAA	411 tyr thr lys lys val pro TAC ACC AAG AAA GTA CCC	58	471 asp arg val thr lys cys GAC AGA GTC ACC AAA TGC	501 glu phe asn ala glu thr phe thr GAG TIT AAT GCT GAA ACA TTC ACC
	231 val	261 ala GCC	291 818 GCC	321 910 GAG	351 1ys AAG	381 val GTG	41 tyr 1AC	441 pro CCT	471 839 CAC	501 91u GAG

5		(1790)	leu ter TTA TAA CATCACATTTAAAAG (1883)	AAATGAAGATCAAAAGCTTATTCATCTGTTTTTCTTTTTCGTTGGTGAAAGCCAACACCCTGTCTAAAAAACATAAATTTCTTTAA (2002)
		560 1ys AAG	¥¢	¥
		540 ala thr lys glu gin leu lys ala val met asp asp phe ala ala phe val glu lys cys cys lys GCA ACA AAA GAG CAA GCT GTT ATG GAT GAT TTC GCT GCT TTT GTA GAG AAG TGC TGC AAG	ter 11AA	1011
		5% 5% 5%	ACAT	AATT
10		lys	CA10	CATA
10		g g Ge	₹	AAAA
		CTA CTA	leu TTA	CTAA
		를 는 - 1 등	41 y	CTGT
	•	818 CCT	3 1	CYCC
15	•	\$ 55 6CT	918 CCC	Ϋ́
		bhe 110	ala	WC
		550 850 GAT	570 ala giu giu giy iya iya leu vai ala ala ser gin ala ala leu giy leu ter GCC GAG GAG GGT AAA AAA CTT GTT GCT GCA AGT CAA GCT GCC TTA GGC TTA TAA	GTGT
		S S	ser AGT	CTTC
		met ATG	\$ \ C \	7770
20		4a1	al a	TCTT
		ala GCT	val	III
		1ys AA	leu CTT	ITCTG
		1e.	1ys AA	ITC
25		g a	1ys AAA	ZT X
23		g a	al y	MAAC
		17 .	glu GAG	XY C
		S&O thr ACA	570 g1u GAG	Š
		ala 60	al a	WAT
30		lys AAG	phe	¥6
		579 273	567 cys 160	. Q
		lys AG	thr	ter ter
		žts Oč	glu	5 5
	•	val lys his lys pro GTG AAA CAC AAG CCC	567 p lys glu thr cys phe a T AG GAG ACC TGC TTT C	ter CATCTCAGCTACCATGAGAAAGA
35		val	asp asp GAC GAT	CCTA
		leu CTC	asp GAC	TCAG
		531 914 646	set ata cct	CATC

TCATTTTGCCTCTTTTCTCTGTGCTTCAATTAAAAAATGGAAAGAATCTAA.... 20AA (2078)

- 10. A nucleotide sequence according to any of claims 6 to 9, in essentially pure form.
- 11. A DNA transfer vector comprising a nucleotide sequence as defined in claim 5.
- 5 12. A DNA transfer vector according to claim 11, transferred to and replicated in a micro-organism.
 - 13. A DNA transfer vector according to claim 12, which is a plasmid.
- 14. A DNA transfer vector according to claim 13,10 wherein the plasmid is pBR322 or YEp6.
 - 15. A process for preparing human serum albumin, which comprises culturing a micro-organism according to claim 5.
 - 16. A DNA transfer vector according to any of
- 15 claims 12 to 14, or a process according to claim 15, wherein the micro-organism is a bacterium or yeast.
 - 17. A vector or process according to claim 16, wherein the bacterium or yeast is <u>E. coli</u> or <u>Saccharomyces</u> <u>cerevisiae</u>.

3.√ 3.√ H Hind III 586 term TAA Sst | 531/2 493/4 Tag I 419/0 450/1 Hinc II Mbo II 472/3 406/7 382 Mbo 13 **pHA36** 36. Kilobases 325/6 Mbo II o: Hinf 1 269/0 ထ Pst 1 (3611) Hps II (3548) 182/3 Tag l Θ. (3611) Hpa II (3548) pHA206 બ 57 Hind 1 Psi 1 Mbo'll Mbo 11 (3611) 16/7 31 Hpa II (3658)

Restriction Endonuclease Map of Human Serum Albumin cDNA Clones